Towards Improving OSS Products Selection– Matching Selectors and OSS Communities Perspectives

Claudia Ayala[1], Daniela Cruzes[2], Xavier Franch[1], Reidar Conradi[2]

1 Technical University of Catalunya
UPC-Campus Nord (Omega), 08034 Barcelona, Spain
{cayala, franch}@essi.upc.edu
2 Norwegian University of Science and Technology
NTNU-Gløshaugen, Trondheim, Norway.
{dcruzes, conradi}@idi.ntnu.no

Abstract. Adopting OSS products is becoming an economical and strategic need for today organizations. A fundamental part of successful adoption is the informed selection of OSS products that best fit the organization needs. One of the main current problems hampering OSS selection is the vast amount of unstructured, incomplete, evolvable and widespread information about OSS products that highly increases the risks of taking a wrong decision as well as the likelihood to deter the use of an OSS product. In this paper, based on marketing research principles, we aim to inform and provide evidence to OSS communities that help them to envisage improvements on their information rendering strategies to satisfy industrial OSS selectors' needs. Our results are from the matching between the informational needs of 23 OSS selectors from diverse software-intensive organizations, and the in-depth study of 9 OSS communities of different sizes and domains. The results evidenced specific areas of improvement. It is a first step to raise the awareness of OSS communities and researchers on areas that are required to improve the industrial OSS selection practice.

1 Introduction

Nowadays, the use of Open Source Software (OSS) provided by OSS communities is revolutionizing the software industry [1]. The fact that OSS products are freely available has influenced not only their significant adoption, but also the way that software is developed and commercialized [10]. Thus, fostering OSS adoption has been recognized as a crucial task for progressing towards improvements in a great variety of application areas [21].

The potential advantages of adopting OSS greatly depend on the ability to select the most suitable product for the task at hand [4]. Improper selection of an OSS product may result in wrong strategic decisions with subsequent economic loss and adverse effects on the business processes of the organizations [13].

In recent years there has been a plethora of proposals aimed to support software products selection, usually suggesting sets of evaluation criteria to evaluate and decide the most suitable alternative(s) (see [15], [18], [13] for comprehensive surveys). However, a recent survey about industrial OSS selection practices [2] shows that these proposals have not been greatly adopted in the industrial practice. Instead, in order to face time-to-market demands and reducing the potential risks, selectors (i.e., the person(s) in charge of the selection process) just base most of their decisions on their experience and tend to limit the use of OSS products to those that are already known and used by the development team. While the value of experience is important, the fact that it is currently considered as the most influential factor for selecting components is at the same time hampering the adoption and fully exploitation of the potential benefits of the high variety of OSS products in the marketplace. Furthermore, such study evidenced that one of the key problems is that selectors are struggling not only with the current diversity of OSS products available in the marketplace, but also with the great deal of widespread, incomplete, heterogeneous, and unstructured information describing each of them (e.g., formal/informal documentation, tutorials, comments in forums, internal experiences) that makes difficult to face a suitable selection process under time-to-market pressures [3], [4], [14]. In addition, the study emphasized that the main source for gathering OSS product information is the OSS community website.

In this context, in order to contribute to enable suitable OSS selection processes we need to envisage more pragmatic approaches than suggesting sets of evaluation criteria (especially when the evidence shows that the data to fill in these criteria is not usually available). Therefore, the goal of this study is to explore the following research questions:

RQ1. *How much of the information required by OSS selectors for performing a suitable selection process is actually provided by OSS communities?*

RQ2: *Are there OSS community characteristics that seem to influence its level of readiness for supporting OSS selection?*

With RQ1, we want to investigate the gap between the information "provided" by OSS communities on their OSS community websites and the information required by OSS selectors to perform an informed selection. RQ2 aims to explore whether some OSS community characteristic(s) seem to affect its readiness (i.e., the degree that the community covers the needs of OSS selectors). By answering these research questions, this paper aims to inform and provide evidence to OSS communities that help them to envisage improvements on their information rendering strategies. It is a first step to raise their awareness on areas that are required to improve the OSS selection industrial practice. This paper reports our results from a pilot study about the matching of 9 OSS communities vs. the needs of 23 industrial OSS selectors.

The rest of the paper is organized as follows: Section 2 provides a brief background of the marketing principles that emphasize the importance of dealing with information rendering aspects and their influence on OSS products selection. Section 3 describes the methodological approach followed to perform the study. Section 4 presents the results obtained from the study, while Section 5 provides a

discussion of main findings. Threats to validity are presented in Section 6. Section 7 summarizes the conclusions and future work.

2 Background

OSS research has largely ignored one interesting aspect that is becoming crucial for OSS projects: first-impression management [7]. Impression management theory refers to the process by which individuals or organizations try to control or manage the impressions that others form of them [23]. Due to the ever-increasing amount of information available on the Internet and the need to make quick choices among competing alternatives, first-impression management has been adopted as one of the main theoretical lenses in marketing literature. Choi et al [7] demonstrated that the OSS community website plays a critical role in attracting developers and users to the community. The mature status of well-known OSS projects likely attracts users given their greater activity and vitality. Furthermore some OSS products have become de facto standards. However this pathway is unavailable for most of the OSS projects and those newly initiated projects that struggle to attract users and contributors [6].

In the context of OSS selection if an OSS project is poorly presented and potential selectors feel that the community does not invest much care in providing the needed information for selection, they might formulate negative opinions about the project and fail to consider it as a candidate even if it might represent a promising alternative. Thus, poor first impressions not only impact on the rate of potential users in the short run, they can also produce negative externalities for the project in the long run. For instance, they might lose the synergies derived from collaborating with companies, such as greater project activity, higher user's base and popularity.

Literature exist on the assessment of factors that might attract developers to participate in OSS projects in order to sustain the vitality of the community [5], [7], [9], [12]. However, as far as we know, there are not empirical studies that consider the needs of industrial OSS selectors as a way to improve first impression management.

Consolidated works from the marketing research have developed relevant tactics to help to influence first impression in a positive way [24]. We think that dealing with first impression management is important for OSS research for two main reasons: 1) OSS community contributors typically join OSS projects by first becoming selectors, subsequently as users and then evolving into contributors [7], therefore first impression management is a potential way to attract OSS potential users. 2) Marketing strategies are becoming crucial to pose OSS products into the marketplace given that nowadays the OSS phenomenon has evolved into a more commercially viable form, where both volunteers and commercial organizations collaborate in its ongoing development [1].

One of the initial grounds of first-impression management is to explore the needs of the potential users and to envisage the improvement tactics. Therefore, this study was designed to explore how the needs of OSS selectors are covered by OSS communities in order to raise observations that may serve as a departing point to envisage suitable improvement tactics.

3 The Study

The study performed in this work was exploratory and aimed to investigate the research questions introduced above. Our main research strategy consisted on the in-depth study of nine OSS community-based projects and exploring how these communities covered the informational needs of 23 industrial OSS selectors.

3.1 Sampling

The target population of the study was OSS community projects. Since the variety of OSS projects is quite wide not only regarding domain and size, but also regarding activity and popularity, we approached a stratified random sampling for improving its representativeness as well as the analysis of the results. We used the Ohloh.net directory as the reference directory for selecting OSS projects. We choose Ohloh as it is one of the largest and up-to-date OSS directories available, and has been widely used to create historical reports about the changing demographics of OSS projects.

From the 437,982 OSS projects referenced in Ohloh by February 2010, we ordered them with respect to their number of users and downloads. Then, we divided such a list into three equal parts (that were considered as stratums). Subsequently, we randomly selected 3 projects from each stratum. Table 1 summarizes the projects that fall into each stratum and provides a brief description of each of them.

Table 1. Stratified random sampling

Stratum	Name	Description
1	Agilo for Scrum	It is one of the most widely used Scrum tools, offering many features to support Scrum and software development teams.
	Joomla	It is an award-winning content management system (CMS), which enables to build Web sites and powerful online applications.
	Subclipse	It is an Eclipse Team Provider plug-in providing support for Subversion within the Eclipse IDE.
2	Gimp	GIMP is an acronym for GNU Image Manipulation Program. It is a freely distributed program for such tasks as photo retouching, image composition and image authoring.
	GNU Grub	GRUB is the GRand Unified Bootloader for GNU
	IpTables	iptables is the user space command line program used to configure the Linux 2.4.x and 2.6.x IPv4 packet filtering ruleset. It is targeted towards system administrators.
3	Fluent NHibernate	A fluent API for simplifying the entity mapping of NHibernate. Add compile time safety, testability, and improved readability to NHibernate projects.
	MediaCoder	It is a free universal batch media transcoder, which integrates most popular audio/video codecs and tools into an all-in-one solution. New codecs and tools are added in constantly as well as support for new devices.
	StatusNet	StatusNet (formerly Laconica) is a microblogging service. Users post short (140 character) notices that are broadcast to their friends and fans using the Web, RSS, or instant messages.

3.2 Data Collection Instrument

In order to assess the OSS projects in a homogeneous way, we developed a data collection instrument based on the survey reported in [2]. This survey provides data about the information that is required in order to perform an informed OSS selection. 23 OSS selectors from 20 small and medium organizations in Spain, Norway and Luxembourg participated in the survey. It consisted on semi-structured interviews that were recorded in audio and then transcribed to text. We had access to the raw data from the respondents of such study (audio and text documents). From the responses of the selectors that participated in the study we elicited a total of 85 informational needs (i.e., specific information that they referred as needed for making informed decisions). We arranged similar answers using content analysis [16] and grouped them in 21 informational needs. Our primary goal was to assess if the information required by the selectors was provided by the OSS communities.

In order to improve the quality of the data collection instrument, it was pre-tested with three researchers. Consequently, we decided to arrange the 21 informational needs in categories and subcategories that provided a more understandable, structured and informative way of collecting them. For instance, the informational need *Time of the product in the market* was grouped in the sub-category *History of the Product* which at its time was grouped in the category *To ensure technological stability and evolution of the OSS product and its provider*. This arrangement demonstrated to provide researchers with a better understanding of the informational needs and their contexts; it therefore enhanced the information gathering process. As a result, the 21 informational needs originally gathered were grouped into 8 categories and 3 sub-categories. The data collection instrument also gather information as: whether the informational need was provided or not by the OSS community; where the information was found, the time required to skim the OSS community website to find the information, and further comments from the researchers that performed and/or reviewed the OSS community. As it can be noted, our intention was not only to explore if the informational needs were available but also to have a first impression about how it was advertised, and how difficult it was to extract it. The resulting data collection instrument is shown in Table 2 and it also provides an example to illustrate the kind of information that was gathered.

3.3 Study Procedures and Data Analysis

Each OSS project was assessed using the data collection instrument introduced above. Two different researchers were in charge on assessing each OSS community website. Subsequently, they discussed and agreed the observations. Once all OSS projects were explored and reviewed, the whole research team held discussion meetings to analyze the data and consolidate the results.

Table 2. Data collection instrument. Example of the assessment of the Agilo for Scrum project

	Category/ Subcategory/ Informational Need	Results of the exploration		
		Status	Where	Further comments
1	Compliance with client's functional requirements			
	List of main functional requirements of the OSS project.	√	(1)	The list of features provided seems quite comprehensive. If further information is required, there is an email available. We asked for further information about the product and our request was quickly processed.
2	To ensure technological stability and evolution of the OSS product and its provider			
	Is it a commercial firm leading the community?	√		The community is lead by a single company called Agile42
	List of companies/organizations collaborating in the community.	×		There is no information about any other company participating in the community.
	History of the product.			
	Time of the product in the market.	×		I navigated through the wiki and could not find this information
	Versions of the product available.	×		
3	Evidence of successful OSS product usage			
	Number of registered users	×		
	List of companies using the product	√	(2)	There is documentation about success stories in companies as: ASDIS, eBuddy, be2, Ericsson, DHD24, Hypoport, Princenton Finantial systems and RES software among others.
	Number of downloads	×		
	Ratings and comments from users	√		In the webpage there are just some textual comments from users as: *"The Agile project approach allowed be2 the ability to monitor the project's performance every two weeks, and to evaluate the performance and quality. agile42 did a perfect job in training and coaching distributed Scrum teams based in both Germany and Armenia for this project. agile42 has been an indispensable link for a successful company transformation."* **Dave Sharrock, Director IT, be2 S.à.r.l.** There is a link http://userstories.com/products/8-agilo-for-scrum that provides comments from users (but this link is sponsored by a webpage called userstories.com). There is also a section called "User's Wishes" to collect potential improvements that the user consider useful….
4	Ease of OSS product integration			
	Interoperability issues			
	List of software system and subsystems required to ensure the correct functioning of the product.	√		Windows/MAC/Linux, andTrac 0.11. The information was widespread and it was not easy to gather it.
	Hardware requirements	×		
	Suitability of Code			
	Well-commented code	√		Very well commented
	Programming language	√		Python
	Availability of support			
5	Free services	√	(3)	Very basic ones most of their services are not free. They have a commercial license where they provide further professional services. By free they offer a blog, google groups for commenting things about the product, and a Wiki.
	Non-Free services	√	(4)	Personalized services. In addition there is an improved version of the product that is not OSS
6	Availability of Tests Results			
	Tests done by the OSS community	×		
	Tests done by an external party	×		
7	Licensing terms			
	Availability of detailed information about the licensing terms and explicitly state if they are listed by the OSS initiative	√	(5)	They have 2 licensing schemas. One that complies with the Apache Software License, and another that is non OSS (offering an improved version of the product)
8	Availability of documentation			
	Documentation for final users	√	(6)	The documentation quality seems acceptable
	Documentation for developers	√	(7)	The documentation for integrators is very basic and scarce
	Available languages of the documentation	☑		Only English
General comments: The page is more oriented to business (the community is lead by a company called Agile42). A wiki is provided to report bugs and to inform about possible further involvement with the community as contributors.				
Name of the researchers: TP + CA **Date of the assessment: 11/2010 – reviewed 02/2011** **Mean time required for skimming the webpage: 4 hours**				

(1) http://www.agile42.com/cms/pages/features/; (2) http://www.agile42.com/cms/pages/references/; (3) https://dev.agile42.com/wiki http://groups.google.com/group/agilo/topics http://agile 42.com/cms/blog/; (4) http://agile42.com/cms/pages/support/; (5) http://agile42.com/cms/pages/agilo/; (6) http://agile42.com/cms/pages/agilo-documentation/; (7) https://dev.agile42.com/wiki/agilo/dev

4 Results of the Study

Results are grouped in two subsections according to the research questions introduced above.

Table 3 shows a summary of the results from the analysis of the 9 OSS communities. In order to provide insights of the coverage of each OSS community to the selectors' needs, we assigned relative weights to each category of the data collection instrument as shown at the right side of each category in Table 3. Such assignment was based on the number of selectors' responses grouped into the category. For instance, the category *Compliance with client's functional requirements* had 18 similar responses; therefore its relative weight with respect to the 85 selector's responses resulted in 21.18%. The category *Availability and quality of the documentation* grouped 3 responses; therefore its weight was 3.53%. Based on such weights we calculated the percentage of coverage of each community to the categories of the data collection instrument. The last row of Table 3 shows the final coverage of each OSS community to the OSS selectors needs. These weights allow us to summarize our findings and provide useful insights to the reader to identify and understand the categories where there is a higher need of improvements.

4.1 How much of the information required by selectors is provided by OSS communities?

We found that the most important informational need belonging to the category *Compliance with client's functional requirements* was covered by all the analyzed projects. All of them show (with diverse levels of detail) a list of features of the OSS product.

The information required *To ensure technological stability and evolution of the OSS product and its provider* was poorly covered by most of the studied OSS communities. In addition, the coverage of informational needs belonging to this category was very diverse (see Standard Deviation in last column of Table 3). Most of the analyzed communities failed to clarify the kind of involvement of commercial firms. While in some cases it was clear that the leader of the community was a commercial firm and that several companies were also collaborating in the community under diverse schemas (coding, sponsoring, donating, etc.), in some other cases this information was not clear. The case of MediaCoder was outstanding as the project has radically changed its OSS nature by a purely commercial approach. At this respect, we found controversial comments in Ohloh claiming that MediaCoder should not be listed therein anymore mainly because the source code is not actually available and this violates one of the principles of OSS [19]. Other projects that did not offer clear information about the involvement of companies were FluentNHibernate and StatusNet, our observations regarding these projects led us to realize that such lack of clarity might come from the fact that these

communities are currently in the process of defining a new business strategy by establishing commercial entities for making business around the products (e.g., selling expert support). Furthermore, basic informational needs as *Time of the product in the market* and *Versions of the product available* were not provided by several communities, especially those with a commercial orientation.

The informational needs grouped in the category *Evidence of Successful OSS product usage* were the ones that most communities failed to cover. None of the studied projects covered all the informational needs belonging to this category. Any of the OSS projects offered information about the number of downloads. Only two communities stated the number of registered users. Five projects stated a list of companies that have successfully used the product, and just one project offered comments from users of the product.

The informational needs belonging to the category *Ease of OSS integration* were mostly covered by the studied projects. The only informational need that was not successfully covered by most projects was related to the *Hardware requirements* needed to ensure the correct functioning of the OSS product. MediaCoder also failed to provide *well-commented* code and *programming language* (this is again due to the fact that it does not provide the source code of the product).

Regarding the category *Availability of support*, in all projects it was explicitly stated whether they provide non-free support, while free support was commonly characterized by wikis, email lists, IRC channels, and forums.

Any project offered information about *Availability of test results*.

Most projects, excepting two (MediaCoder and StatusNet) offered clear information regarding *Licensing terms*. As mentioned above, these two projects were facing a business model change and therefore their licensing schemas were not clearly stated. Finally, regarding the *Availability of documentation*, almost all projects offered documentation for final users and for developers and most of them offered a variety of languages.

Summarizing, we found that the analyzed OSS projects cover the selectors' needs in a diverse degree. Such coverage ranges from 44.96% to 80.89%. Further discussions are provided in section 5.

Table 3. Summary of results

Category/Subcategory/Informational Need	Agilo for Scrum	Joomla	Subclipse	Gimp	GNU Grub	IpTables	Fluent NHibernate	Media Coder	StatusNet	Standard Deviation
Compliance with client's functional requirements (21.18%)	21.18	21.18	21.18	21.18	21.18	21.18	21.18	21.18	21.18	0
List of main functional requirements of the OSS project.	✓	✓	✓	✓	✓	✓	✓	✓	✓	
To ensure technological stability and evolution of the OSS product and its provider (21.18%)	5.30	15.89	10.59	21.18	21.18	21.18	10.59	5.30	0	7.28
Is the project governed by a commercial firm or by the community?	✓ commercial	✓ community	✓ commercial	✓ GNU	✓ GNU	✓ community	Not clear	✓ commercial	Not clear	
List of companies/organizations collaborating in the community (others than the leader)	×	✓	×	✓	✓	✓	×	×	×	
History of the product.										
Time of the product in the market.	×	✓	×	✓	✓	✓	✓	×	×	
Versions of the product available.	×	×	✓	✓	✓	✓	✓	×	×	
Evidence of successful OSS product usage (12.94%)	6.47	6.47	3.24	3.24	3.24	0	0	0	3.24	2.53
Number of registered users	×	✓	✓	×	×	×	×	×	×	
List of companies using the product	✓	✓	×	✓	✓	×	×	×	✓	
Number of downloads	×	×	×	×	×	×	×	×	×	
Ratings and comments from users	✓	×	×	×	×	×	×	×	×	
Ease of OSS product integration (12.94%)	9.71	9.71	12.94	12.94	9.71	9.71	6.47	9.71	9,71	1.94
Interoperability issues										
List of software system and subsystems required to ensure the correct functioning of the product.	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Hardware requirements	×	×	×	✓	✓	×	×	✓	×	
Suitability of Code										
Well-commented code	✓	✓	✓	✓	✓	✓	✓	×	✓	
Programming language	✓	✓	✓	✓	✓	✓	✓	×	✓	
Availability of support (11.76%)	11.76	11.76	11.76	11.76	11.76	11.76	11.76	11.76	11.76	0
Free services	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Non-Free services	✓	✓	✓	×	×	×	✓	✓	✓	
Availability of test results (9.41%)	0	0	0	0	0	0	0	0	0	0
Tests done by the OSS community	×	×	×	×	×	×	×	×	×	
Tests done by an external party	×	×	×	×	×	×	×	×	×	
Licensing terms (7.06%)	7.06	7.06	7.06	7.06	7.06	7.06	7.06	0	0	2.77
Availability of detailed information about the licensing terms and explicitly state if they are listed by the OSS initiative	✓	✓	✓	✓	✓	✓	✓	×	Not clear	
Availability of documentation (3.53%)	2.6	3.53	2.60	3.53	3.53	3.53	2.60	0.25	3.53	1.06
Documentation for final users	✓	✓	✓	✓	✓	✓	✓	×	✓	
Documentation for integrators	✓	✓	✓	✓	✓	✓	✓	×	✓	
Available languages of the documentation	☑ Only English	✓ Several	☑ Only English	✓ Several	✓ Several	✓ Several	☑ Only English	☑ Only English	✓ Several	
Mean time spent by the two researchers for skimming the portal (hrs):	4:00	4:50	3:05	3:02	3:03	3.50	2:25	3:00	2:43	
Resulting percentage of coverage of the portal to the needs of selectors:	64.07	75.59	66.13	80.89	80.89	74.42	62.90	44.96	49.41	12.13

4.2 Are there OSS project characteristics that influence its level of readiness for supporting selection?

The assessment of the 9 OSS community projects leads us to state some observations regarding characteristics that might affect the information rendering aspects of OSS communities and therefore their readiness for supporting selection. The most relevant ones suggest that two interrelated characteristics seem to affect the information rendering aspects of OSS projects: the involvement of commercial firms and the stratums that the projects belong to.

We observed substantial differences among the 3 Stratums. Surprisingly, all OSS projects from Stratum 1 have a close involvement of commercial firms in the community. This finding is in line with the results from the study reported in [5] that evidenced that firms coordinate, develop code for, or provide libraries to one third of the 300 most active OSS projects in SourceForge. Projects from Stratum 2 did not have commercial firms leading the projects, instead they referred to volunteer-based communities that fully adhered to the Free Software Foundation (FSF) and two of them (Gimp and Grub) were part of the GNU project that advocate for the "free software" philosophy. Projects from the Stratum 3 also show a high involvement of commercial firms. While MediaCoder is currently a purely commercial project, Fluent NHibernate and StatusNet are facing a transition stage for becoming business-oriented OSS communities. These facts are in line with the "commercialization" of OSS predicted by [10].

The involvement of firms in the OSS communities seems to influence their aesthetic appearance and information rendering aspects. For instance, Agilo for Scrum is an OSS project entirely governed and led by the company Agile42, and so its website is more oriented to business (i.e., selling services around the product) than to promote the involvement of potential contributors to the community. In the case of the Joomla!, the involvement of firms seems to be quite different as even if firms are quite involved in the project, the project is governed by the community. Thus, its website reflects a strong interest to promote resources for consolidating the community and attract contributors. It also offers several schemas for companies and organizations to participate in the community (i.e., donations, selling services around the product, merchandizing). The website of Subclipse is led by the company CollabNet and the provided resources are more oriented to final users (i.e., instructions on how to install the plugging) than to contribute to the community. Other examples are FluentNHibernate and StatusNet that are currently approaching business oriented models and are also improving the aesthetic appearance of their portals. Therefore, we suggest that: as higher the involvement of commercial firms is, the lesser seems to be the attention paid to promote the involvement of potential contributors in the community.

Communities without commercial firms involved shared several commonalities. It seems that they are mostly aimed to provide technical resources to strengthen the developers' community than aesthetic and attractive resources. This coincides with some studies that emphasize that some OSS projects mostly leaded by community programmers often value substance over form and some exhibit an antipathy for marketing and public relations work [11]. In all cases, these OSS projects provide mailing lists, forums and wikis aimed to enable the collaboration among the members.

5 Discussion of Results

The previous section aimed to present a comprehensive view of the results. This section aims at emphasizing and discussing the most important findings and observations.

One of the main difficulties we faced in our OSS projects assessment was that the information—even if it was sometimes available—was not directly accessible. We had to browse the project website and explore among help files, manuals, or even demos. This fact increased the time spent on skimming the portal to find the information and definitely rules out any possibility of trying to automate the search for the information as previously stated by [3].

Even if we found that the involvement of commercial firms and the stratums seem to have a significant influence on information rendering aspects, the coverage of selectors' needs varies from project to project.

Furthermore, OSS communities are not aware of the importance of making some information available. We observed that some informational needs are actually known by OSS community but are not explicitly provided. For instance, most of the analyzed projects did not explicitly offer information about the *Number or registered users* in the community, *Time of the product in the market*, or *List of companies using the product* (if any). So, we hope that the results provided here help to raise the awareness of the importance of providing such information. In addition, there are categories that seem to be almost always provided (i.e., *Compliance with client's functional requirements* and *Availability of support*) while there are others that are not fully covered (e.g., *To ensure technological stability and evolution of the OSS product and its provider*) or are not covered by any of the studied project as *Availability of test results*. Providing such evidence is important to envisage the corresponding improvement strategies and increase the competitive advantage of the OSS products.

6 Threats to Validity

This section discusses the threats to validity of our study in terms of construct, internal, and external validity, as suggested by [20] and [22]. It furthermore emphasizes the corresponding strategies used to deal with these threats.

6.1 Construct Validity

Regarding construct validity, our study was supported by 2 main principles: rigorous planning of the study, and the establishment of protocols and instruments for data collection and data analysis. The data collection instrument was carefully designed taking into account the informational needs of selectors elicited from a semi-

structured interview further reported in [2], as detailed in section 3.2. This allows us to focus the study on the information that is really needed by the industrial OSS selection practice. In addition, the data collection instrument was pre-tested and enhanced by creating categories and subcategories for grouping informational needs (as detailed in section 3.2). This allows us to improve its understandability and therefore to improve the data gathering process.

6.2 Internal Validity

Regarding internal validity, we tried hard to envisage and harmonize the data gathering and the subsequent data analysis strategies. With respect to the data gathering strategy, we took relevant decisions for approaching a better understanding of the availability of the information for covering the selectors' needs. One of the most relevant decisions was to avoid the non-deterministic factors inherent to the OSS selection processes. These non-deterministic factors refer to contextual issues that greatly affect the OSS selection decision. For instance, even if the OSS project provides a list of functional characteristics of the product, it might happen that such a list is not detailed enough for the context of the selection project and the selector have to face such a lack of detail by testing the product himself or by looking for further information in forums, email lists, etc. The strategies for facing (or not) such lack of information depend on the amount of time and resources that a company is willing and able to invest in the selection process [8]. Therefore, to avoid such potential issues we decided to focus our observations just on whether the informational need was covered or not and on how the information was provided.

In addition, we decided that two different researchers independently faced the assessment of each OSS community projects using the data collection instrument. Subsequently they discussed their results in order to agree and merge them. This helps us to deal with the potential subjectivity of the assessment of each researcher. Furthermore, it is important to mention that the researchers participating in the study are impartial parties and do not have any kind of involvement with any of the OSS communities analyzed. In this sense, we consider that there is no any intentional bias regarding the data gathered.

6.3 External Validity

Regarding external validity, it is important to highlight that the character of our study is exploratory, and hence we did not aim to make universal generalizations beyond the studied setting, but also provide some observations that might serve as a departing point for further investigations and improvements. Having this in mind, we discuss some mitigation strategies used in the study.

One of the main threats of external validity of the study is that we approached a small set of OSS community projects and these projects might not represent the whole variety of OSS projects. We tried to mitigate any possible bias related to this

by having a stratified random sampling so that the studied OSS communities were diverse regarding size, application domain, popularity and success.

The informational needs used as a base to decide the informational coverage of OSS projects were elicited from industrial OSS selectors. While extracting such needs from the industrial practice is a good point to strengthen the external validity of our observations, we are aware that eliciting such needs from 23 selectors might not represent all real needs. However, we think that such results are useful to have a first approximation to the problem and might serve as a basis to envisage future studies.

Finally, other issues that might affect the presented results (especially the time spent skimming the OSS projects for finding and understanding the information) are:

a) As mentioned above, the assessment of each OSS project was performed using a strategy that avoids the non-deterministic nature of the OSS selection processes (i.e., just capturing whether the informational need was covered or not and further observations about how it was provided). At this respect, we are aware that in the industrial OSS selection practice the complexity and time for gathering the OSS products information is actually higher. Thus, we would like to stand out that the metrics (weights and time for assessing each OSS project) provided in Table 3 are just intending to offer insights of the coverage of each OSS community to the selectors' needs. In any case these metrics are aimed to be representative of the assessment of OSS projects for any specific OSS selection process.

b) The researchers in charge of analyzing the OSS communities were not experts in any of the approached domains. So, we may say that the performance of researchers that performed the data gathered process would be more similar to "novice selectors" than experienced selectors that might perform better in finding and analyzing the OSS communities.

7 Conclusions and Future Work

This study presents our results of exploring the current gap between the "required" information needed by 23 industrial OSS selectors for making informed decisions and the information "provided" by 9 OSS communities. The obtained results would contribute to research and practice: a) by informing OSS communities about information rendering aspects that could be improved to attract industrial users. b) by informing OSS selection researchers about informational limitations that might help them to calibrate their OSS selection proposals. This is important because there is no point in defining theoretical criteria and measures for selecting OSS products if the data they rest upon is not available.

Our future work focus on complementing the results from the study reported here with further information that allows OSS communities to elaborate their tactics of improvement based on the selectors' feedback [24].

8 References

[1] Ayala, C.P., Cruzes, D., Hauge, Ø.; Conradi, R. Five Facts on the Adoption of Open Source Software. IEEE Software. March-April 2011, pp. 95-99.

[2] Ayala, C., Hauge, Ø., Conradi, R., Franch, X., Li, J. Selection of Third Party Software in Off-The-Shelf-Based Software Development - An Interview Study with Industrial Practitioners. The Journal of Systems & Software 84 (2011), pp. 620-637.

[3] Bertoa, M., Troya, J.M.,Vallecillo, A., 2006. A Survey on the Quality Information Provided by Software Component Vendors. Journal of Systems and Software 79, 427-439.

[4] Boeg, J., 2006. Certifying Software Component Attributes. IEEE Software, 23(3), 74-81.

[5] Bonaccorsi, A., Lorenzi, D., Merito, M., Rossi, C. Business firms' engagement in community projects. Empirical evidence and further developments of the research, in Proceedings of the First International Workshop on Emerging Trends in FLOSS Research and Development (FLOSS 2007), May 21th, Minneapolis, USA, IEEE Computer Society, 2007 pp. 1–5, doi:10.1109/floss.2007.3

[6] Capiluppi, A., Lago, P, Morisio, M. "Evidences in the Evolution of OS Projects through Changelog Analyses," Proc. 3rd IEEE Workshop Open Source Software Eng. (WOSSE 03), ICSE, 2003, pp. 10–24.

[7] Choi, N., Chengalur-Smith, I.,Whitmore, A. Managing First Impressions of New Open Source Software Projects. IEEE Software . November-December 2010. pp., 73-77.

[8] Cruz, D., Wieland, T., Ziegler, A., 2006. Evaluation Criteria for Free/Open Source Software Products Based on Project Analysis. Software Process: Improvement and Practice, 11(2), 107-122.

[9] Denner, C.S., Pearson, J., Kon, F., Attractiveness of Free and Open Source Projects. The European Conference on Information Systems (ECIS) 2010. ISBN: 978-0-620-47172-5

[10] Fitzgerald, B. The Transformation of Open Source Software. MIS Quarterly, 30(3), 2006.

[11] Fogel, K. Producing Open Source Software: How to Run a Successful Free Software Project, O'Reilly, 2006.

[12] Hauge, Ø., Ayala, C.P., Conradi, R.: Adoption of open source software in software-intensive organizations - A systematic literature review. Information & Software Technology 52(11): 1133-1154 (2010).

[13] Jadhav, A.S., Sonar, R.M., 2009. Evaluating and Selecting Software Packages: A review. Information and Software Technology, 51(3), 555-563.

[14] Li, J., Conradi, R., Bunse, C., Torchiano, M., Slyngstad, O.P.N., Morisio, M., 2009. Development with Off-The-Shelf Components: 10 Facts. IEEE Software. 26(2), 80–87.

[15] Li, J., Conradi, R., Slyngstad, O.P.N., Torchiano, M., Morisio, M., Bunse, C., 2008. A State-of-the-Practice Survey of Risk Management in Development with Off-the-Shelf Software Components. IEEE Transactions on Software Engineering, 34(2), 271-286.

[16] Krippendorff, A., 1980. Content Analysis. Sage Publications, London.

[17] Mahmood, S., Lai, R., Kim, Y.S., 2007. Survey of Component-Based Software Development. IET Software, 1 (2), 57-66.

[18] Merilinna, J., Matinlassi, M., 2006. State of the Art and Practice of Open-Source Component Integration. Proceedings of the 32nd EUROMICRO Conference on Software Engineering and Advanced Applications. IEEE Computer Society, pp. 170-177.

[19] Open Source Initiative http://www.opensource.org/

[20] Robson, C., 2002. Real World Research: A Resource for Social Scientists and Practitioner-researchers. Second Edition. Blackwell Publishers Inc.

[21] Simmons, G.L., Dillon, T.S., 2006. Towards an Ontology for Open Source Software Development". In IFIP International Federation for Information Processing, Volume 203, Open Source Systems. Damiani, E., Fitzgeralg, B., Scacchi, W., Scotto, M., Succi, G., (Eds). Boston:Springer, pp 65-75.

[22] Wohlin, C., Runeson, P., Host, M., Ohlsson, M.C., Regnell, B., Wesslen, A., 2000. Experimentation in Software Engineering - An Introduction. Kluwer Academic Publishers.

[23] Winter, S.J., Saunders, C., Hart, P. "Electronic Window Dressing: Impression Management with Websites," European J. Information Systems, vol. 12, 2003, pp. 309–322.

[24] Wolfe, E., Bies, R. Impression Management in the Feedback-Seeking Process: A Literature Review and Research Agenda. *Academy of Management Review*. 1991. Vol 15, No. 3, pp. 522-541.